$$S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

Aa $\Sigma\sigma$

Descriptive Statistics

Shape: Bell (Normal) Skewed, uniform
clusters, gaps, outliers.

Center: Mean, mode, median

Spread:

$\underbrace{\text{Range, IQR}}_{\text{measures spread around median}}$, $\underbrace{\text{Standard Deviation}}_{\text{measures spread around mean.}}$

Range: min to max. IQR: middle 50%

Review of notation

x = value

$\sum x$ = Sum of all values

n = Sample size

Population ← Sample

↓ ↓

Parameter ← Statistic

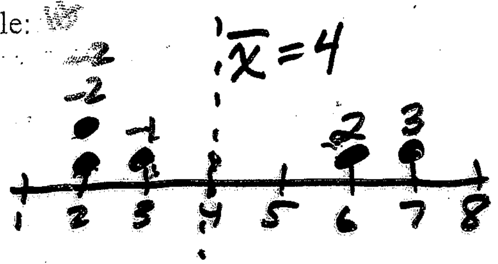
$\mu = \frac{\Sigma x}{N}$ mean $\qquad \bar{x} = \frac{\Sigma x}{n}$ mean

What is Standard Deviation?

$$\underbrace{\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}} \qquad s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}}_{\text{Standard Deviation.}}$$

We start by understanding Deviation =
Difference between a data value and the mean of the data set.

For example:



$-2 + (-2) + (-1) + 2 + 3 = 0$

We want a Single # to represent the spread of a data set that takes into account All of the Deviations for that data set. We call this number the Standard Deviation.

Definition of Standard Deviation:

→ Tells us, On Average, how far away values fall in either direction of the mean.

→ Measure of variation or dispersion of all values from the mean.

Why?

- measure of center does not tell us how wide a graph is

Measuring variability (Spread) in numerical data allows a more complete description than just stating a central measure.

How? – The Theory behind the equation.

When you want to measure the spread of data, it is typical to start by finding the mean. Next we find the Deviations, or direct distance, from each data value to the mean. Then we find the Average of those ~~values.~~

$$\sqrt{\frac{\sum(\overset{\text{values}}{x}-\overset{\text{mean}}{\bar{x}})^2}{n-1}}$$

The problem: All Deviations add to 0.

2 possible solutions:

~~1) Absolute Value~~

2) Square each # ← makes each Deviation positive

How do we find the average and why do we divide by n-1? ← Degrees of freedom

$\frac{\sum x}{n-1}$ 20

5 5 5 5

9 1 6, $\boxed{4}$

Independent

dependent

If don't we underestimate the true s.d.

Last: Since we squared the deviations in the beginning, we now have to take the square root to find the final value of our standard deviation. The final equations:

to compensate for squaring we need to take the square root.

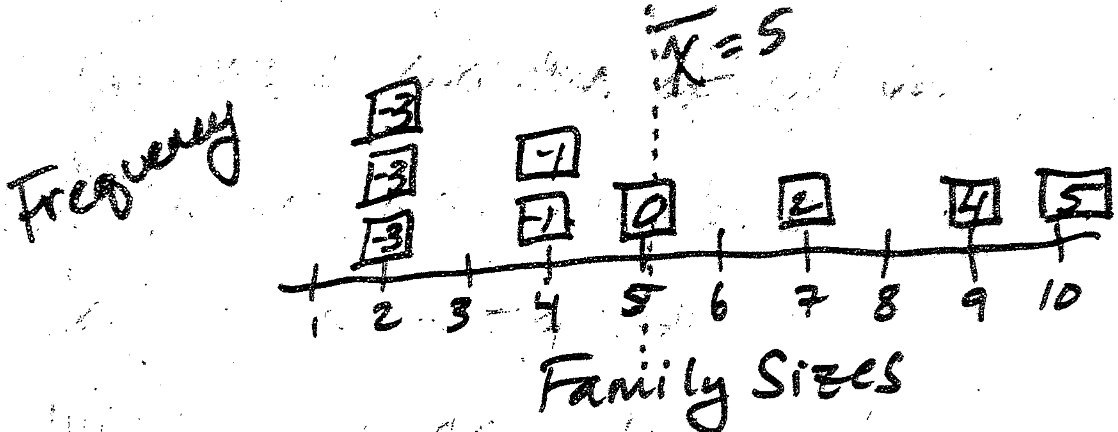
Example.

Calculator

Interpretation.

Set #2

Putting the Equation together with an example:



Step 1.) Find $\bar{x}$ (the mean) of the sample data set.

$\bar{x} = 5$

Step 2.) Find each deviation from the mean: $x - \bar{x}$

1VarsStat.

$-3, -3, -3, -1, -1, 0, 2, 4, 5$

Step 3.) Square each deviation to make it positive: $(x-\bar{x})^2$

$9, 9, 9, 1, 1, 0, 4, 16, 25$

Step 4.) Sum all of the squared deviations: $\sum(x-\bar{x})^2$

74

Step 5.) Divide by the degrees of freedom: $\dfrac{\sum(x-\bar{x})^2}{n-1} = \dfrac{74}{8} = 9.25$

$n-1$

$9-1=8$

Step 6.) Take the square root:

$\sqrt{9.25} = 3.04$ $\qquad \sqrt{\dfrac{\sum(x-\bar{x})^2}{n-1}}$

So $s = \sqrt{\dfrac{\sum(x-\bar{x})^2}{n-1}} = 3.04$

What does this mean?

The average family size data values are 3.04 units from the mean.

Standard Deviation.

Some general ideas to keep in mind:

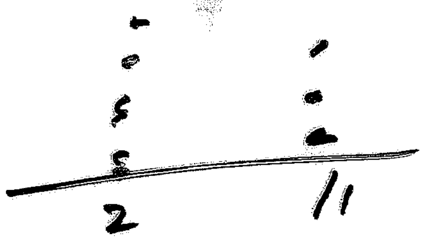
) Measure of variation of all values from the mean.

2.) Always a positive – can be 0.

3.) Larger values – greater variation.

4.) Will increase dramatically by outliers.

Population $\sigma = 5.4$ ← Sample $s = 4.85$



$s = 4.3$